



# Exploration du site web scholasticon.fr : une application de la methode SyMoGIH (Système modulaire de gestion de l'information historique)

Francesco Beretta

## ► To cite this version:

Francesco Beretta. Exploration du site web scholasticon.fr : une application de la methode SyMoGIH (Système modulaire de gestion de l'information historique). Cabouret, Bernadette; Demotz, François. La prosopographie au service des sciences sociales, CEROR, pp.289-310, 2014, 978-2-904974-48-9. halshs-00988861

**HAL Id: halshs-00988861**

**<https://shs.hal.science/halshs-00988861>**

Submitted on 12 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Exploration du site web *scholasticon.fr* : une application de la méthode SyMoGIH (Système modulaire de gestion de l'information historique)

Le site *scholasticon.fr* contient plus de 2200 notices bio-bibliographiques consacrées à des philosophes et théologiens de l'époque moderne. Il a été créé en 1999 par Jacob Schmutz, maître de conférences à l'Université Paris-Sorbonne, spécialiste de philosophie médiévale et de la Renaissance. Lors de la rédaction de sa thèse consacrée à « La querelle des possibles. Recherches philosophiques et textuelles sur la métaphysique jésuite espagnole, 1540-1767 » (soutenue en 2003), il avait réalisé non seulement une présentation bio-bibliographique complète des vingt principaux auteurs sur lesquels portait sa recherche, mais encore des notices courtes pour la plupart des auteurs cités.

Afin de valoriser ces ressources et de les mettre à la disposition de la communauté des chercheurs, J. Schmutz a créé un site web statique ayant « pour vocation d'offrir la meilleure documentation possible sur la culture scolastique moderne, c'est-à-dire la philosophie et la théologie des professeurs ayant enseigné en Europe et dans le Nouveau Monde, dans les collèges, couvents et universités, pour la période 1500-1800 »<sup>1</sup>. Le critère qui décide de l'insertion d'un acteur dans la population 'Scholasticon' est donc le fait d'avoir enseigné l'une de ces deux disciplines à l'époque en question. Outre les notices bio-bibliographiques de philosophes et théologiens, le site comprend une importante bibliographie de la production historiographique récente consacrée à Francisco Suárez (1548-1617). Dans les années suivantes, *scholasticon.fr* a été alimenté grâce aux contributions de plusieurs collègues et correspondants<sup>2</sup> et il est ainsi devenu un site de référence dans ce domaine d'études car on y trouve de nombreuses informations concernant des auteurs peu connus et difficilement accessibles par ailleurs.

### **Evolution de *scholasticon.fr* vers un système d'information prosopographique**

Le projet a toutefois rencontré quelques difficultés, communes à ce type de réalisation. D'une part, il y a le problème de la pérennisation de l'hébergement du site et de sa mise à jour technologique. Grâce à une collaboration entre le Pôle ressources informatiques de l'Institut des sciences de l'homme de Lyon (ISH) et le Pôle méthodes du

1 "http://scholasticon.fr/Presentation/index\_fr.php". Tous les sites web indiqués ont été consultés en février 2013.

2 Voir la liste des contributeurs : "http://scholasticon.fr/Presentation/Contact\_fr.php".

Laboratoire de recherche historique Rhône-Alpes (LARHRA), le site a été refait en 2008 et est hébergé sur un serveur de l'ISH. D'autre part, un site web collaboratif de ce type demande la mise en place d'un système d'information dynamique adapté aux ambitions du projet : on souhaite en effet disposer d'une navigation souple entre les biographies des auteurs, leur références bibliographiques, les institutions d'enseignement, leurs ordres religieux d'appartenance, etc.

Pour apporter une solution efficace à ces problématiques, une base de données collaborative, appelée Base d'hébergement de projets (BHP), a été mise en place au LARHRA dès 2008 dans le cadre du projet Système modulaire de gestion de l'information historique (SyMoGIH)<sup>3</sup>. Elle accueille les données d'une trentaine de chercheurs et étudiants, ainsi que celles de cinq projets de recherche. Les données peuvent être regroupées par projet et publiées sur une site dédié, comme le montre l'exemple du projet de bio-bibliographie des entrepreneurs français « Patrons de France »<sup>4</sup>, dont le site web publie des données stockées dans la BHP. Le site permet de naviguer facilement de la notice d'un patron à celle d'une ou plusieurs institutions d'appartenance, ou de représenter sur une carte ses lieux d'activité. En effet, la dimension géographique a été intégrée dès le départ au projet SyMoGIH dans le but de mettre en place un système d'information collaboratif géo-historique<sup>5</sup>.

La question se pose donc de la transformation possible du site *scholasticon.fr* en vue d'un hébergement dans la BHP et de la publication des informations sous une nouvelle forme, analogue à celle du site *patronsdefrance.fr*. Cette évolution permettrait de gérer plus aisément l'alimentation du site par différents contributeurs et de lui offrir une nouvelle dynamique en termes de navigation et de ressources mises à la disposition des chercheurs et du public. Surtout, cette évolution permettrait de 'libérer' la richesse en informations que contient le site sous forme de notices-texte, en transformant ces informations non-structurées en données munies d'une qualification sémantique. Elles deviendraient ainsi exploitables pour une analyse de type prosopographique.

Grâce au site *scholasticon.fr*, nous disposons en effet d'une population de plus de 2200 acteurs, réunis grâce à un critère précis : le fait d'avoir enseigné la philosophie ou la théologie au cours de la période 1500-1800. Il faut toutefois relever que le site englobe des notices bio-bibliographiques alimentées au fil des recherches personnelles de J. Schmutz et des autres contributeurs, avec la finalité de mettre à la disposition du public des informations utiles mais sans aucun projet systématique de récolte de données.

---

3 ["http://www.symogih.org"](http://www.symogih.org).

4 ["http://www.patronsdefrance.fr"](http://www.patronsdefrance.fr).

5 Voir le portail de ressources SIG ["http://www.geo-larhra.org"](http://www.geo-larhra.org).

Certaines notices sont très riches, d'autres se réduisent à une seule ligne.

Sans vouloir reprocher aucunement cet état de fait aux producteurs du site, dont le mérite est reconnu par la communauté des spécialistes, on peut se demander quel est le contenu effectif de ces notices en termes d'informations utilisables dans la perspective de la démarche prosopographique. Il s'agit donc, d'une part, d'une question de curiosité intellectuelle : y a-t-il des informations directement disponibles dans ces textes, permettant de mettre en lumière l'un ou l'autre phénomène sociologique ou intellectuel ? Et, d'autre part, la réponse à cette question est la condition préalable à une évaluation du coût effectif de l'opération envisagée, c'est-à-dire transformer *scholasticon.fr* en un système d'information géo-historique dynamique.

Pour répondre à ces questions, nous avons appliqué la sémantique du projet SyMoGIH aux notices disponibles sur le site web afin d'en extraire un premier jeu de données, puis d'explorer ces données grâce à quelques outils logiciels de visualisation et de comptage. Le travail d'extraction des données a été effectué en 2010, sous ma direction et en collaboration avec J. Schmutz, par un étudiant non spécialiste du domaine : Arthur Jacob a parcouru en trois mois les 2128 notices disponibles à l'époque et a introduit dans la BHP les données tirées de ces textes en fonction de certains critères. Dans la suite de mon exposé, je présenterai, d'abord, les principes essentiels de la sémantique du projet SyMoGIH, en expliquant de quelle manière elle a été appliquée pour produire les données à partir des informations disponibles dans les notices de *scholasticon.fr*. Ensuite, je proposerai une exploration de ces données en vue de répondre aux questions posées : les informations contenues dans le site web permettent-elles, en l'état actuel, de leur appliquer la démarche prosopographique ? Quelles sont les conditions pour qu'elles le deviennent ?

## **La méthode SyMoGIH et les éléments essentiels de sa sémantique**

Le projet Système modulaire de gestion de l'information historique (SyMoGIH), développé dès 2008 au sein du Pôle méthodes du LARHRA, est né de la volonté de mettre en place un système collaboratif de gestion de l'information géo-historique permettant de valoriser l'effort de collecte des données, individuel et par projet, en rendant celles-ci interopérables et donc réutilisables pour d'autres recherches. Étant donné l'expérience du LARHRA en matière de bases de données, notre effort s'est concentré sur cette technologie informatique dans le but de mettre en place un modèle des données générique et ouvert, permettant de mutualiser les informations récoltées.

La sémantique de SyMoGIH s'efforce de reproduire informatiquement la méthode de collecte de l'information effectuée par l'historien avec un procédé qui s'inspire de

l'utilisation classique de la fiche Bristol, tout en lui appliquant trois principes : la décomposition poussée de l'information en unités de connaissance atomisées ; la définitions d'objets (acteurs, institutions, lieux, etc.) mis en relation à travers ces unités de connaissance ; la documentation de chaque unité de connaissance par une ou plusieurs sources, en application d'un principe essentiel de la méthode historique : le sourçage de l'information.

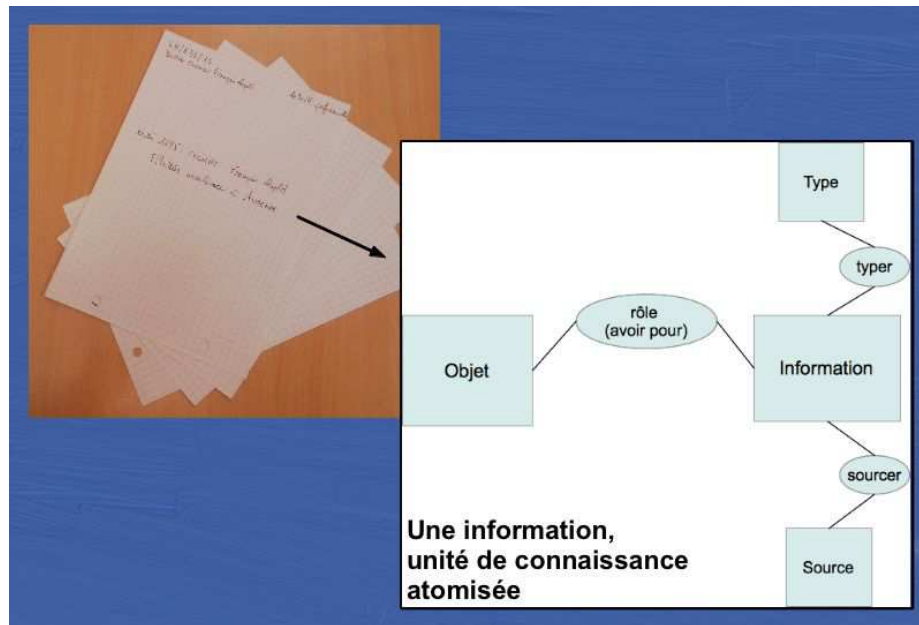


Figure 1. Le principe de SyMoGIH : produire un modèle informatique de la fiche Bristol

L'atomisation de l'information représente la condition de la réutilisation des données pour des recherches différentes de celles qui les ont produites, en leur appliquant d'autres problématiques. D'où l'effort de produire des données primaires (*primary data*) munies d'un caractère le plus objectif possible. Il s'agit, bien entendu, d'un but visé car une objectivité conçue dans l'abstrait n'existe pas : toute donnée résulte nécessairement d'un questionnement, d'un regard précis qui détermine sa production. Toutefois, dans un système d'information géo-historique collaboratif, différentes stratégies permettent de se rapprocher du but visé. Il s'agira de poser au cœur du système un modèle des données le plus générique possible ; de définir des objets et des relations entre eux le plus possible proches de la réalité sociale, même dans leur formulation ; de renvoyer le codage des données, réalisé en fonction d'une problématique, à la phase de l'exploitation tout en se limitant à stocker dans la base de données les relations élémentaires qui subsistent entre les objets ; et de documenter avec précision toute instance concrète du modèle générique,

afin que la sémantique ayant produit les données soit explicite et facilement compréhensible par tous les utilisateurs.

Lorsque le chercheur extrait des connaissances à partir d'un texte, il le fait en règle générale dans deux optiques différentes : soit il veut reproduire le discours de la source, les dates des événements et leur interprétation telles que le document les fournit, même s'il sait que ces dates ou ces interprétations ne sont pas correctes ; soit il s'intéresse à reconstituer ces connaissances comme telles, en affirmant que ces événements se sont effectivement passés à telle date et selon telle version des faits. Dans notre système d'information, nous avons appelé 'contenus' le premier type de connaissances produites par l'historien et 'informations' le second. Si on s'intéresse maintenant au volet des 'informations', voici une version simplifiée du modèle générique :

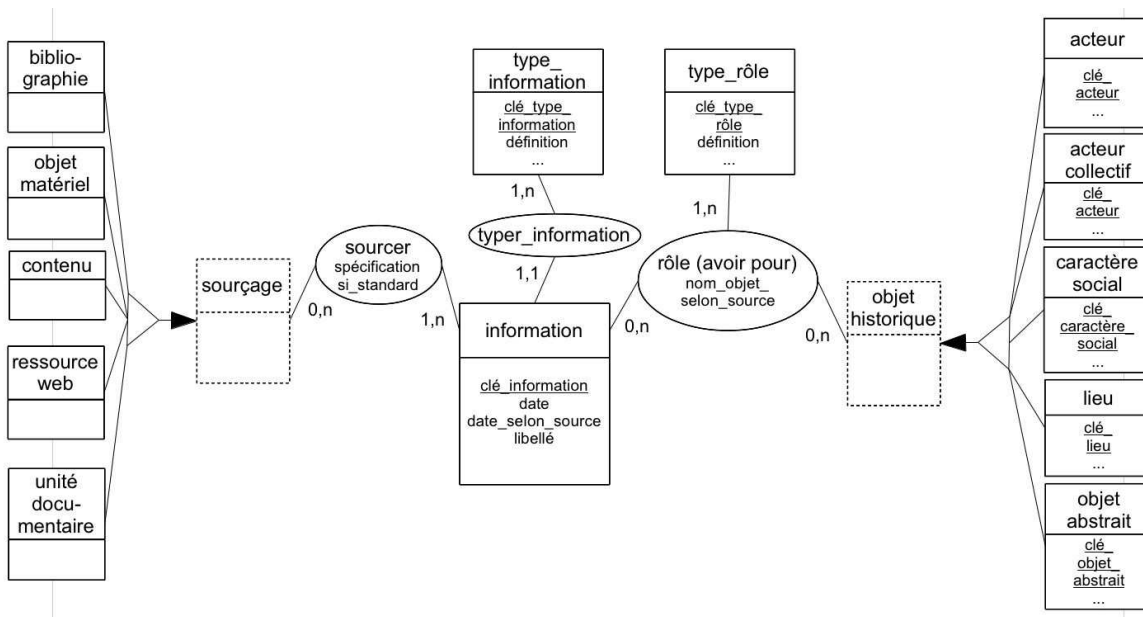


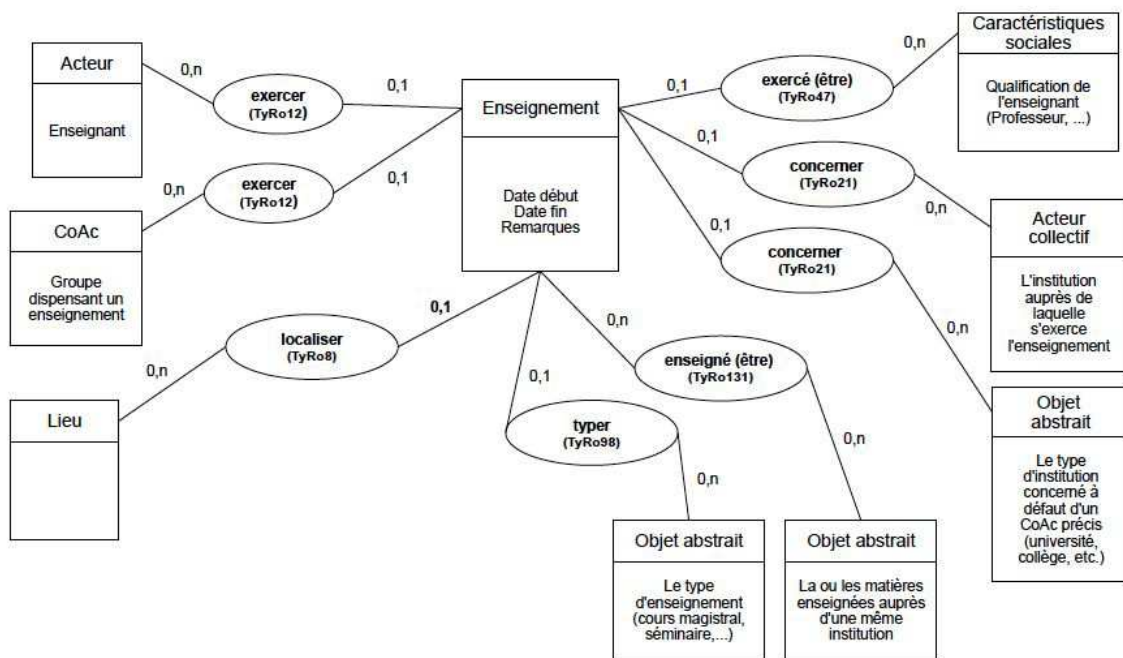
Figure 2. Le modèle générique des 'informations'

Ce modèle générique est exprimé selon le formalisme de la méthode Merise, méthode classique en modélisation des bases de données. Dans le modèle conceptuel de SyMoGIH, une information représente une unité de connaissance atomisée, qui exprime une relation entre objets (acteurs, institutions, concepts, lieux, etc.), chacun de ces objets participant à cette relation avec un rôle propre. L'information possède des propriétés, tels une date, ou un libellé indiquant le sens de l'association entre les objets, et elle est qualifiée par un type qui en spécifie la nature. Par exemple, une information de type

naissance met en relation un acteur avec le lieu de sa naissance et elle a comme propriété une date. De plus, chaque information doit être munie d'au moins une source, la source étant elle-même l'un des objets ou, plus précisément, l'une des classes d'objets définies dans le système.

Ce modèle générique très simple se limite donc, à un premier niveau, à faire état d'associations datées entre objets, munies d'une explication sous forme de texte et d'un type qui spécifie le sens de l'association. À un second niveau, on applique concrètement ces principes à un cas précis, en modélisant différentes instances du modèle générique, dites 'types d'information', qui qualifient le sens des informations et définissent les rôles propres à chaque objet associé. Cette étape s'effectue au fur et à mesure de l'avancement des recherches, en dialogue entre les spécialistes de différents domaines et ayant toujours en vue l'objectivité visée dans la collecte des données primaires. Voici, à titre d'exemple, la modélisation retenue pour une information de type 'enseignement' :

Enseignement (TyIn97)



Alexandre Giandou, Francesco Beretta, 10 mai 2010 © CNRS  
Modifications par F.Beretta 2.6.2010

Figure 3. Modèle conceptuel du type d'information 'enseignement'

Ce modèle conceptuel se limite à décrire la relation subsistant entre la personne

qui enseigne, la matière enseignée, l'institution auprès de laquelle est dispensé l'enseignement, la qualification du grade de l'enseignant, éventuellement son lieu si l'institution n'est pas connue ou si la localisation est différente de celle du siège de l'institution. Ce modèle n'est aucunement absolu, et ne constitue pas la seule manière de traiter un enseignement : il a pour fonction de documenter la manière adoptée pour produire les données, en se limitant à exprimer des relations primaires entre les objets. Si des cas de figure plus complexes se présentent, demandant une décomposition plus fine des relations entre objets, d'autres modèles seront produits, tout en les documentant et en fournissant une explication détaillée de leur articulation avec le premier modèle.

La problématique du chercheur, le questionnement qui dirige sa recherche ne seront pleinement mis en œuvre qu'au moment d'extraire les données grâce à des requêtes qui s'appuient sur la sémantique de la base de données collaborative, documentée par la modélisation des types d'information, et qui utilisent le langage d'interrogation SQL (Structured Query Language). Etant donné la structure très simple du modèle générique, des requêtes de base peuvent être pré-configurées pour aider le chercheur à s'initier à leur utilisation ; ensuite, les requêtes peuvent devenir plus complexes afin de prendre en compte différents paramètres, de regrouper et de coder automatiquement les données, etc. SQL est un langage puissant, qui permet de retravailler et de recomposer les données avant de les exporter vers des logiciels permettant leur visualisation et leur analyse : statistique, analyse des réseaux, systèmes d'information géographique (SIG), etc.

Concernant les connaissances à extraire des notices de *scholasticon.fr*, il a paru opportun d'utiliser la partie 'informations' de la sémantique de SyMoGIH, et non les 'contenus', puisque le genre littéraire des notices biographiques s'efforce de représenter des connaissances condensées et vérifiées : il ne s'agit donc pas, dans ce cas, de reproduire le discours de la source mais de structurer les connaissances exprimées sous forme de notices. Celles-ci ont été relues systématiquement pour en extraire toutes les informations de type 'enseignement', codées selon le modèle indiqué. De plus, on a renseigné, dans la mesure du possible, la naissance des enseignants (comprenant la date et le lieu), leur formation, leur appartenance à une ou plusieurs organisations (en particulier aux ordres religieux) et la géolocalisation des différentes institutions<sup>6</sup>.

Avec cette méthode 9000 informations env. ont été produites pour une population de 2128 acteurs (état en août 2010), auxquelles s'ajoutent 1000 informations env. déjà présentes dans la base ou ajoutées ultérieurement, ce qui représente une valeur moyenne actuelle de 4,7 informations par acteur appartenant à la population 'Scholasticon', avec

---

6 La sémantique propre à ces différents types d'information est consultable sur le site "symogih.org".



une valeur médiane de trois informations. Cet écart entre la moyenne et la médiane résulte du caractère non-systématique des notices, voire des lacunes qu'elles présentent souvent, ce qui fait que pour certains acteurs ont pu être extraites quelques dizaines d'informations, alors que pour d'autres, on dispose de très peu de connaissances. Ajoutons que dans les notices manque l'indication explicite de la source des informations qu'elles contiennent. Par conséquent, dans le sourçage actuel des données de la BHP, on a indiqué qu'elles sont tirées de *scholasticon.fr*, en sachant qu'on peut certes faire confiance aux rédacteurs des notices, mais qu'en même temps, d'après un principe essentiel de la méthode historique appliqué dans la BHP, ces informations ne sont pas susceptibles de vérification et qu'elles ne sont donc pas réutilisables comme telles dans d'autres contextes.

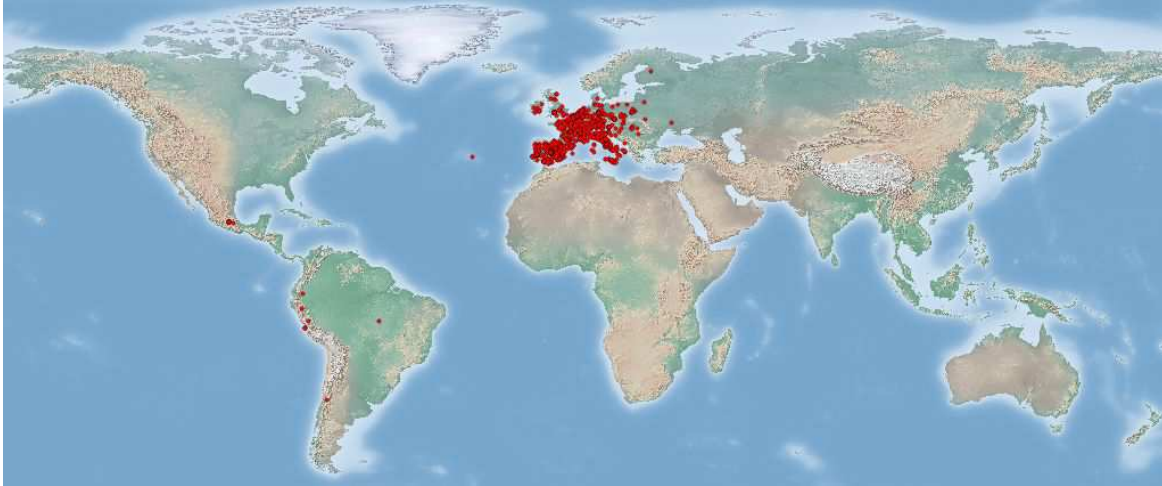
### **Exploration des données issues des notices de *scholasticon.fr***

Grâce aux données produites selon la méthode indiquée, il a été possible de réaliser une première exploration des informations que contient le site web et d'en mettre en évidence les contours et les limites pour une approche de type prosopographique. Je résumerai ici quelques résultats de cette exploration afin d'illustrer l'état actuel des données et d'indiquer quelques pistes susceptibles d'être parcourues pour les compléter. Les différents outils logiciels mobilisés par la suite serviront uniquement à explorer et à visualiser les données, et non à les analyser car, en l'état actuel, ces données ne se prêtent pas à un traitement systématique. Ce constat découle en toute logique de la nature des notices qui, comme nous l'avons indiqué dès le départ, n'ont pas été produites en vue de la récolte systématique d'informations mais au fil des recherches de leurs producteurs.

Que contiennent donc les notices de *scholasticon.fr* en termes de données produites utilisant la sémantique de SyMoGIH ? Nous pouvons commencer ce bref parcours d'exploration par une distribution des naissances dans l'espace, à l'échelle de la planète<sup>7</sup>.

---

7 J'utilise à cette fin le logiciel QuantumGIS, "<http://www.qgis.org/>", en prenant comme fonds de carte un raster tiré du site "<http://www.naturalearthdata.com/>".



*Figure 4. Lieux de naissance de l'ensemble de la population*

On constate sans surprise que la densité des naissances est plus élevée en Europe. On peut aussi relever la localisation des naissances en Amérique latine, se regroupant autour de régions particulières. Ce sera aux spécialistes de ces régions de juger s'il s'agit d'un phénomène significatif ou si ce constat dépend simplement du caractère sélectif des données disponibles. En comptant ensuite les acteurs par lieu de naissance et en représentant les effectifs ainsi calculés, on obtient le résultat suivant :

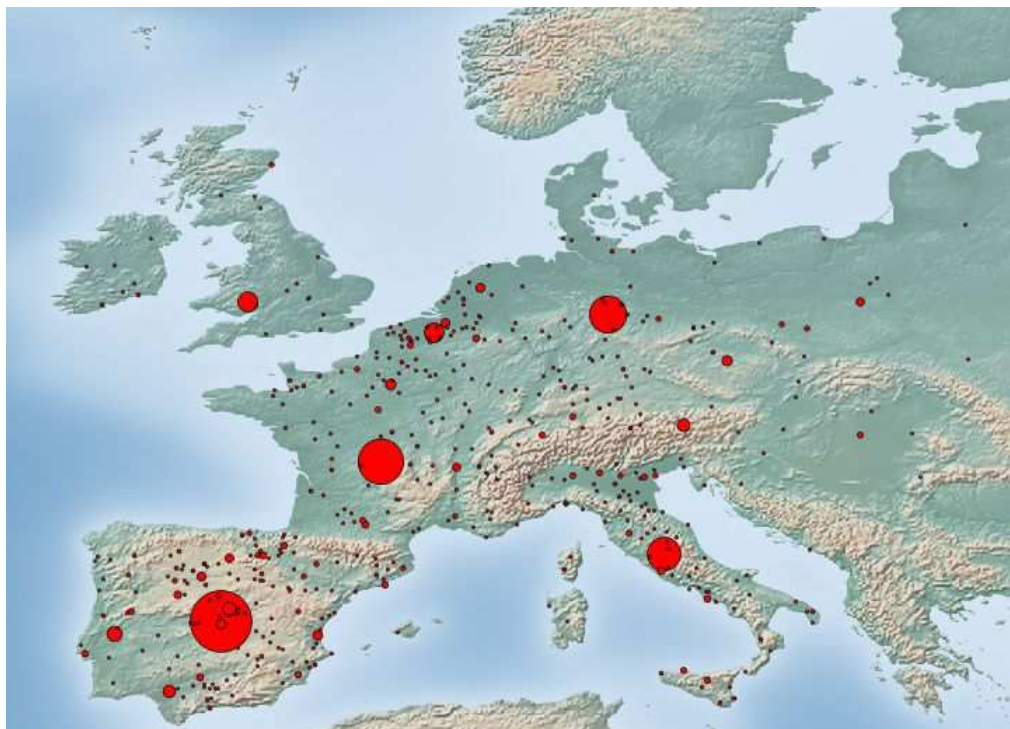


Figure 5. Effectifs par lieu de naissance, en Europe

On sera surpris par les effectifs des points se situant au centre de l'Espagne, de la France, de l'Italie, etc. Ce phénomène s'explique par le fait que, souvent, les notices ne donnent pas d'indication précise concernant le lieu de naissance : on sait qu'il s'agit d'un théologien espagnol, français, etc. Pour disposer d'un lieu de naissance géolocalisé, il a donc fallu créer de points génériques correspondants au centroïde de la région géographique en question. Le même problème se pose par rapport à l'année de naissance, inconnue dans de nombreux cas. Nous avons essayé d'utiliser certains paramètres pour postuler une année de naissance à partir des indications de la notice mais l'imprécision reste importante. Si l'on s'interroge sur la part de population disposant d'un lieu et d'une date de naissance précise, on obtient le résultat suivant :

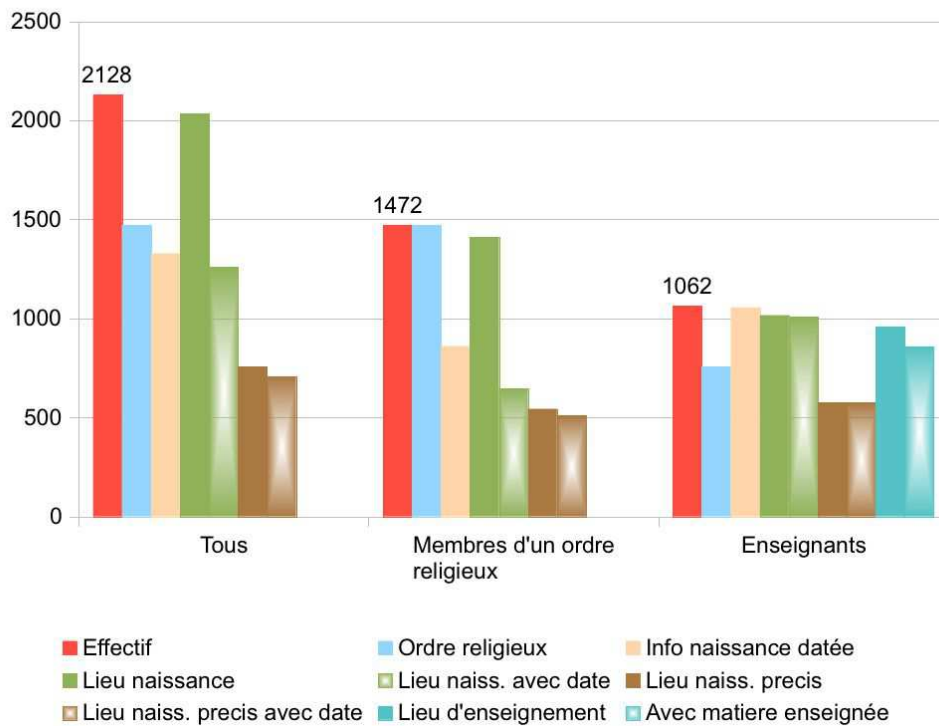


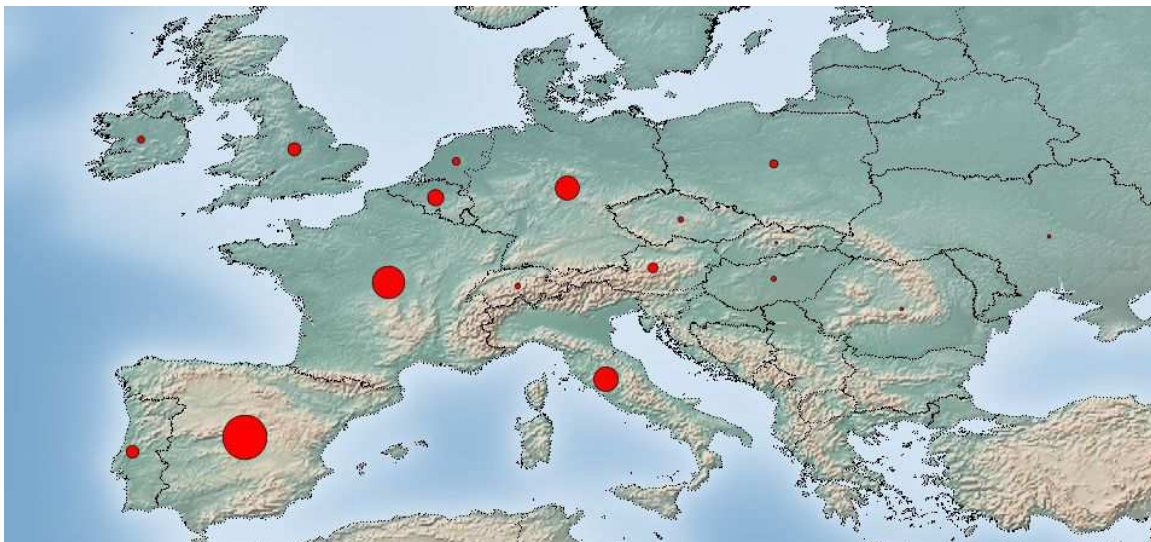
Figure 6. Regroupement synthétique des données disponibles

Les valeurs des effectifs pour l'ensemble de la population (figure 6, partie de gauche) montrent que des indications précises concernant le lieu et la date de naissance sont disponibles uniquement pour un tiers de la population (en brun). En restreignant l'analyse aux membres d'ordres religieux (partie centrale), la proportion ne change pas

significativement. On relèvera, en revanche, que l'appartenance à un ordre religieux est une caractéristique des deux tiers de la population : il s'agit d'un élément important, sur lequel nous reviendrons.

Les données ne présentent une relative homogénéité que si l'on se limite aux acteurs disposant d'une information relative à un enseignement (partie de droite) et, là encore, avec des lacunes. Dans ces cas, l'enseignement peut, en règle générale, être localisé et associé à une matière enseignée : nous allons donc orienter notre exploration autour de la question de la localisation et du contenu des enseignements. Mais il est d'ores et déjà possible de relever que, dans une population définie par rapport au critère de l'enseignement, les éléments manquent pour la moitié des acteurs pour remplir correctement une information de type 'enseignement'. Ce qui montre le caractère souvent imprécis des informations que contiennent les notices. Pour poursuivre notre visualisation en disposant de données relativement homogènes, nous devons donc 'abandonner' la moitié de la population du site web et nous limiter aux données concernant les 1062 'enseignants'.

On peut alors se demander comment regrouper et compter les individus par 'nation'. Cela peut se faire grâce à une analyse spatiale, étant donné que les informations sont géolocalisées, mais à condition de disposer d'une carte des frontières. À ma connaissance, il n'y a que des données spatiales géolocalisées contemporaines qui sont gratuitement accessibles à des fins de recherche<sup>8</sup> :



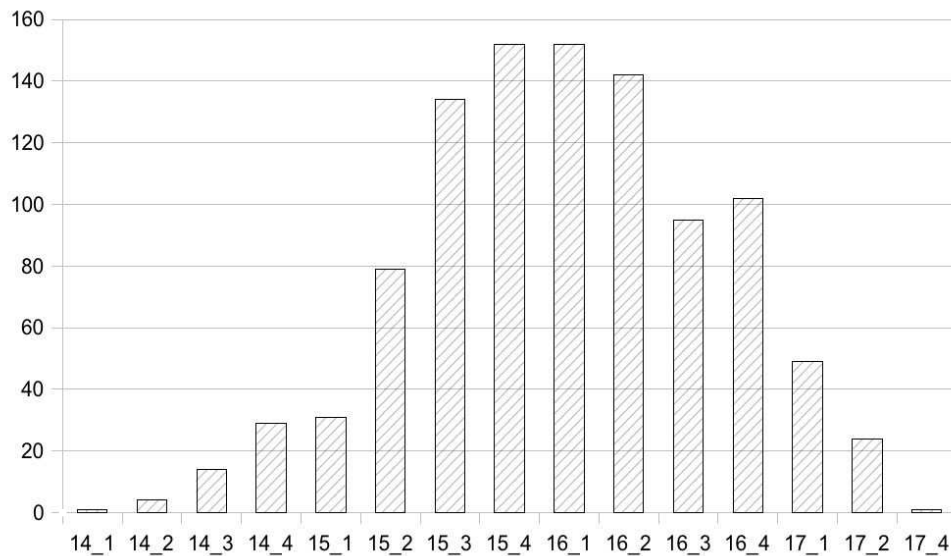
*Figure 7. Effectifs des naissances par pays*

---

8 J'ai utilisé les données du projet Global Administrative Areas (GADM), "<http://www.gadm.org/>".

Or, si cette carte montre bien la prépondérance des acteurs nés en Espagne, elle pêche par son anachronisme car les comptages de l'Europe centrale n'ont pas de sens par rapport à l'Europe de la première modernité. Se pose ainsi la question de l'acquisition de données spatiales plus pertinentes. Seule l'entreprise Euratlas semble disposer à ce jour de données historiques vectorisées et géolocalisées sur les frontières des États pour l'époque moderne. Un inconvénient de ces données est représenté par leurs 'coupes' chronologiques car les frontières sont cartographiées chaque siècle, c'est-à-dire, pour l'époque en question, en 1500, 1600, 1700. Du point de vue de l'historien, ce choix n'est pas très pertinent étant donné les modifications significatives des frontières qui ont été ratifiées lors de différents traités au milieu du 16<sup>e</sup> siècle, ou lors des traités de Westphalie, en 1648.

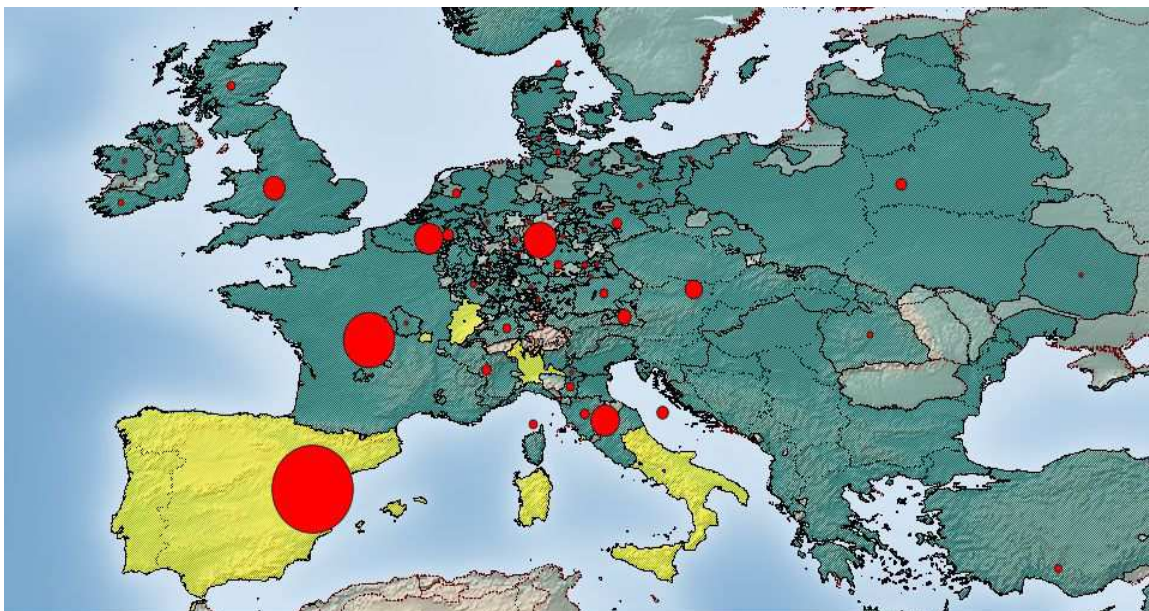
À défaut de mieux, nous avons été amenés à utiliser ces données spatiales et, en vue de choisir la carte la plus adaptée, il a paru opportun de regrouper en générations de 25 ans les acteurs, en fonction de leur année de naissance connue ou postulée, puis d'afficher une distribution de ces classes d'acteurs regroupés par quart de siècle :



*Figure 8. Regroupement des acteurs par 'génération' (périodes de 25 ans)*

L'histogramme montre clairement, et un calcul rapide le confirme, que plus de la moitié des acteurs sont nés entre 1550 et 1650. Ce qui invite à retenir la carte de 1600 pour calculer de la manière historiquement la plus pertinente possible les effectifs des naissances par pays :





*Figure 9. Effectifs des naissances par pays (frontières en 1600 ©EurAtlas 2009)*

Cette carte montre la nette prédominance des sujets du roi d'Espagne (territoires en jaune dans la figure 9), qui sont tous regroupés en un seul point représentant l'effectif le plus important. Aussi, la carte met une nouvelle fois en évidence le problème lié à l'imprécision des données extraites de *scholasticon.fr* car la localisation des naissances par centroïdes des régions géographiques contribue, par exemple, à gonfler le nombre de ressortissants des Etats pontificaux parmi lesquels l'analyse spatiale englobe indument tous les 'Italiens' génériques, en raison du positionnement du centroïde de l'Italie.



*Figure 10. Imprécision des données spatiales et erreurs de comptage*

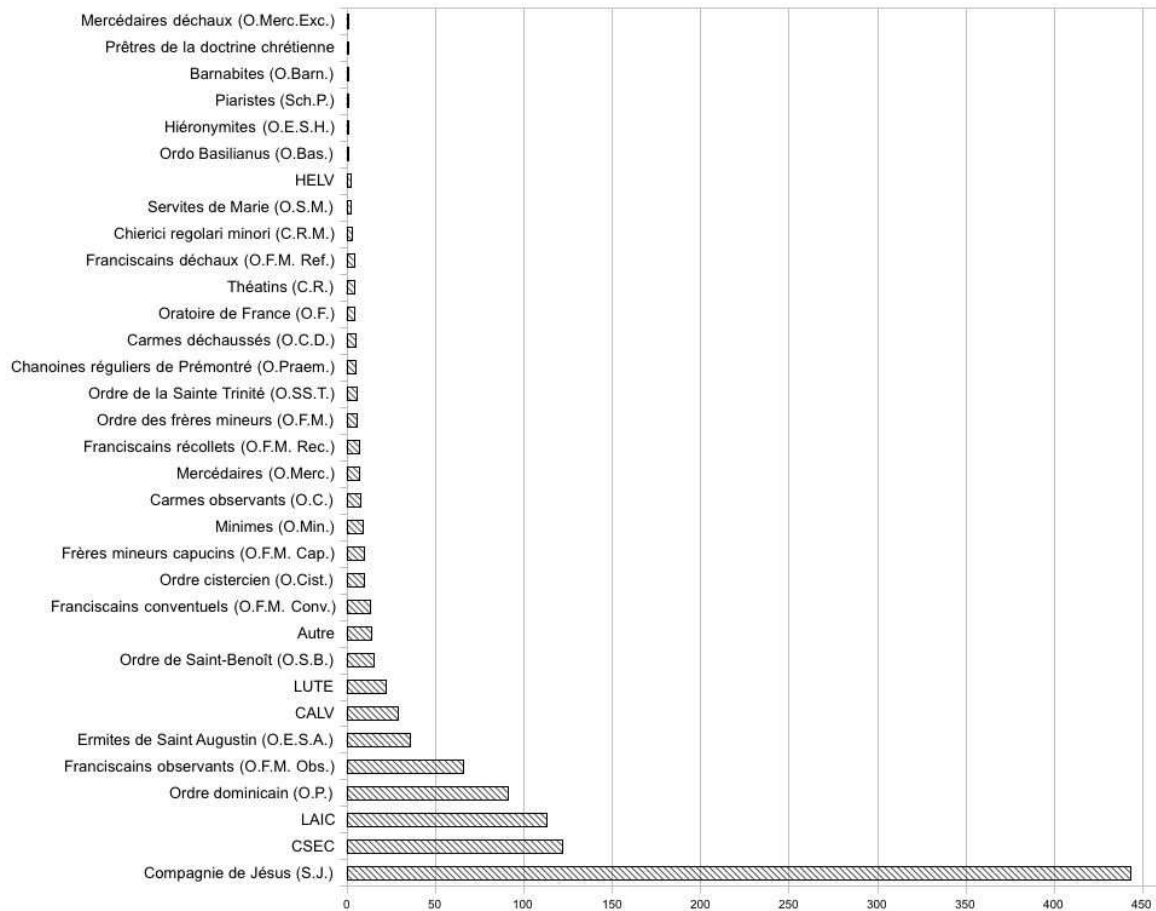
Enfin, est apparu un problème important lié à la qualité des données spatiales. Nous avons constaté des erreurs de comptage en utilisant les données produites par Euratlas (en rouge dans la figure 10) par rapport au jeu de données contemporaines (en jaune), ce dernier étant beaucoup plus précis. Vérification faite, la raison du décalage dans les effectifs apparaît clairement : les acteurs nés en dehors de la ligne des côtes seront 'perdus' dans le décompte, comme c'est le cas pour deux théologiens nés dans la ville de Trapani, à l'ouest de la Sicile. Devant le constat des difficultés qui surgissent suite à l'absence de frontières géolocalisées et historiquement pertinentes, nous avons entamé au sein de SyMoGIH un projet de modélisation et de traitement des données spatiales, qui permettra de réaliser à terme un atlas historique plus adéquat, tant en termes de précision des données spatiales que de gestion de l'évolution des territoires<sup>9</sup>. Étant donné l'ampleur d'un tel projet, nous espérons pouvoir créer un réseau réunissant d'autres institutions ayant déjà produit, du moins pour certaines régions géographiques, ce type de données<sup>10</sup>.

L'analyse générale des données présentée ci-dessus (figure 6) a montré que le fait d'être membre d'un ordre religieux est une caractéristique propre à une large partie de la population de *scholasticon.fr*. En se limitant au groupe des 'enseignants', tel qu'il a été défini, et qu'en regroupant les acteurs en fonction de l'ordre d'appartenance, on obtient la distribution suivante :

---

9 Voir "<http://geo-larhra.ish-lyon.cnrs.fr/?q=atlas-historique>".

10 Entre autres : "<http://www.visionofbritain.org.uk/>"; "<http://www.atlas-europa.de/>"; "<http://www.hgis-germany.de/>"; "[http://www.censusmosaic.org/cgi-bin/index\\_data\\_gis.plx](http://www.censusmosaic.org/cgi-bin/index_data_gis.plx)".



*Figure 11. Distribution de la population par ordre religieux ou 'situation'*

L'histogramme montre d'emblée la nette prépondérance des membres de la Compagnie de Jésus, qui représentent presque 42 % de la population (figure 11). On remarquera également les sigles en majuscules, issus d'un codage des données visant à les rendre plus lisibles. Il a paru judicieux de mettre en évidence, à côté des religieux, la proportion des professeurs laïcs, ou des clercs non réguliers (CSEC), et également celle d'acteurs qui, selon les notices du site web, appartiennent à l'une des confessions réformées, ou dont celle-ci est déductible en fonction du lieu d'enseignement. Il apparaît ainsi que calvinistes, luthériens et zwingliens représentent 5 % de la population d'enseignants. Cet exemple illustre la distinction propre à la sémantique SyMoGIH entre le stockage des données primaires et le codage en fonction d'une problématique de recherche. En effet, il aurait été imprudent de rattacher directement des acteurs à une confession lors de la production des données, alors que cette appartenance n'est pas



certaine.

En termes de qualité des données, d'importantes limites apparaissent du point de vue d'une approche prosopographique. D'une part, il a été parfois délicat, à partir des textes des notices, de décider l'attribution d'un acteur à un ordre plutôt qu'à un autre, en particulier pour les acteurs appartenant à la vaste famille franciscaine. D'autre part, la prépondérance des membres de la Compagnie de Jésus, qui se recoupe pour de nombreux acteurs avec leur origine ibérique, découle de la thèse de doctorat de l'initiateur du site et provoque un déséquilibre important dans la population. Le fait qu'il y ait seulement un peu plus que 10 % de professeurs laïques ne reflète sans doute pas la situation effective de l'époque : il suffit de penser aux centaines de professeurs de philosophie naturelle des universités italiennes. Aussi, l'ouverture à l'Europe de la Réforme apparaît très judicieuse, mais l'effectif de ces professeurs est actuellement trop réduit pour envisager des comparaisons.

Ce déséquilibre dans la constitution de la population apparaît également lorsque l'on visualise, grâce à une boîte à moustaches (*boxplot*)<sup>11</sup>, la distribution de la population en fonction de son année de naissance et de l'appartenance à un ordre religieux ou à sa 'situation' pour les laïcs et les réformés. Les rectangles au milieu de la figure 12 indiquent le volume et la distribution dans le temps de la moitié centrale de chaque sous-population.

---

11 Pour réaliser la boîte à moustache j'ai utilisé le logiciel R, "<http://www.r-project.org/>", alors que pour les histogrammes je me suis servi du logiciel Calc de la suite LibreOffice, "<http://fr.libreoffice.org/>".

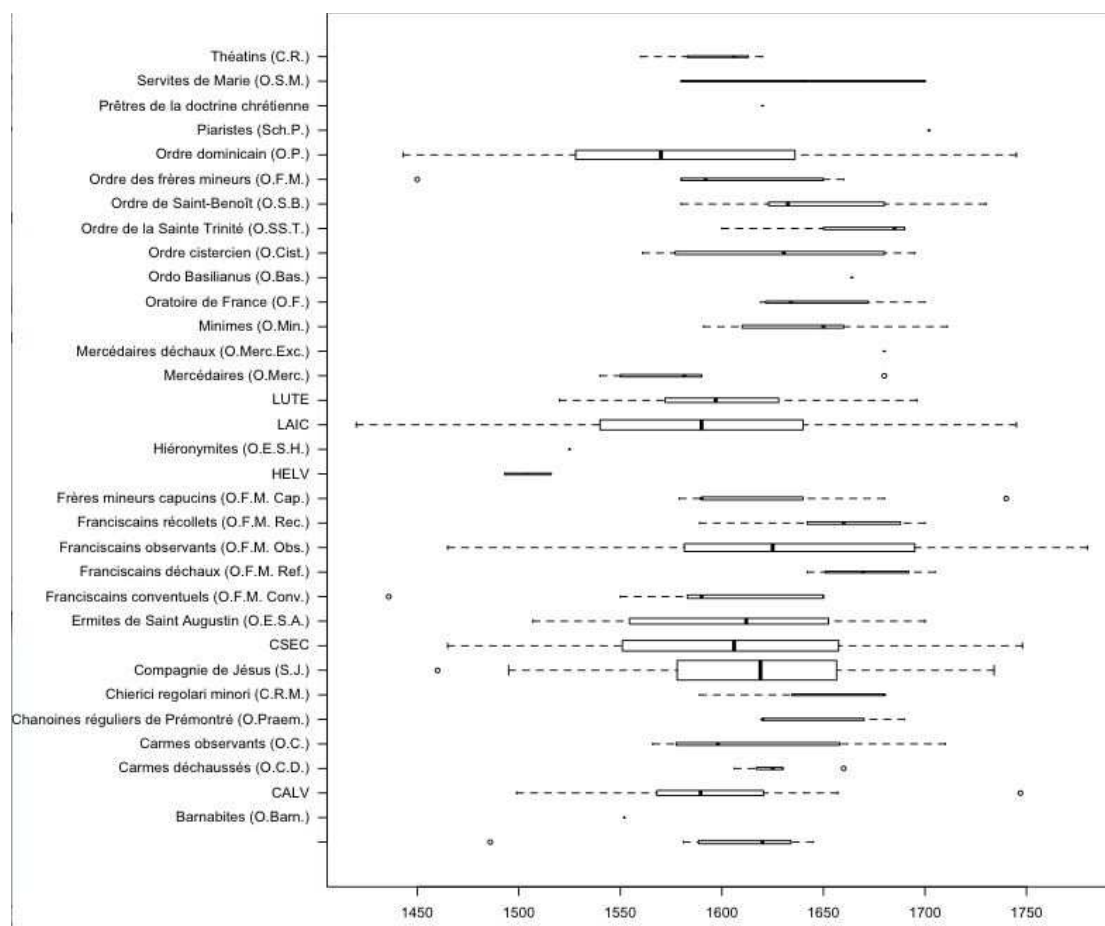
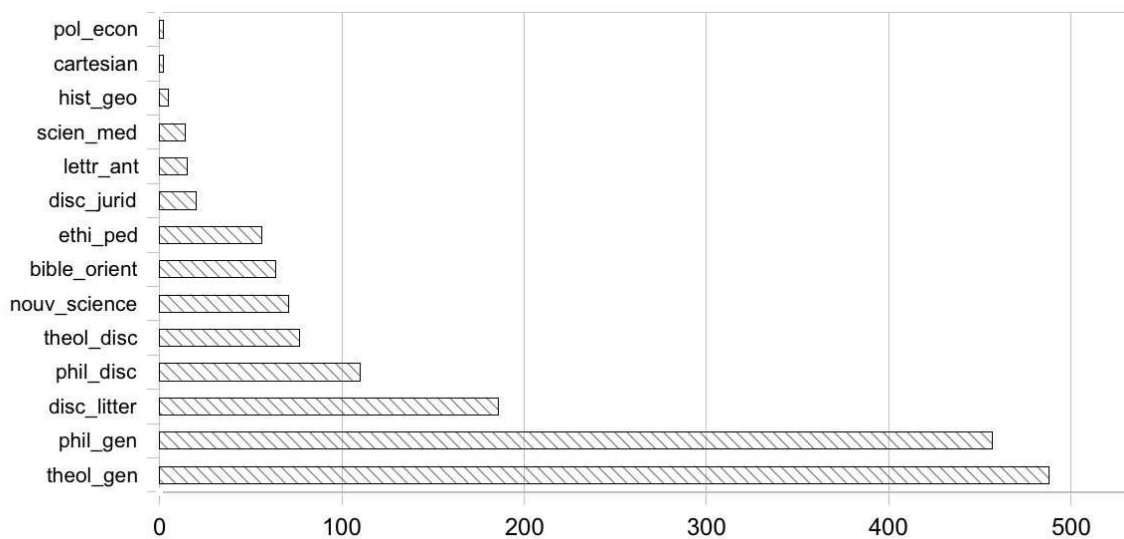


Figure 12. Distribution des acteurs par 'situation' et par année de naissance

Dans ce tableau, qui permet de se faire une assez bonne idée des 'forces' en présence, apparaît clairement un décalage temporel, qui peut s'expliquer certes par des raisons structurelles pour les membres d'institutions apparues seulement au 17<sup>e</sup> siècle, comme l'Oratoire de France, mais qui est bien plus dû au caractère incomplet des informations sur l'Ordre dominicain ou les membres des confessions réformées.

Des difficultés analogues surgissent lorsque l'on s'intéresse aux disciplines enseignées : un travail important de réflexion sur les objets correspondants a été nécessaire en vue de la production des données. En effet, les notices du site web sont souvent trop imprécises pour décider du rattachement à l'une ou à l'autre discipline ou mélangent dans leur formulation le statut de l'enseignement (par ex. « chaire de vêpres ») avec la matière enseignée, ce qui n'est pas aisé à distinguer pour un non-spécialiste. De plus, un regroupement et un codage des disciplines se sont imposés pour rendre visibles

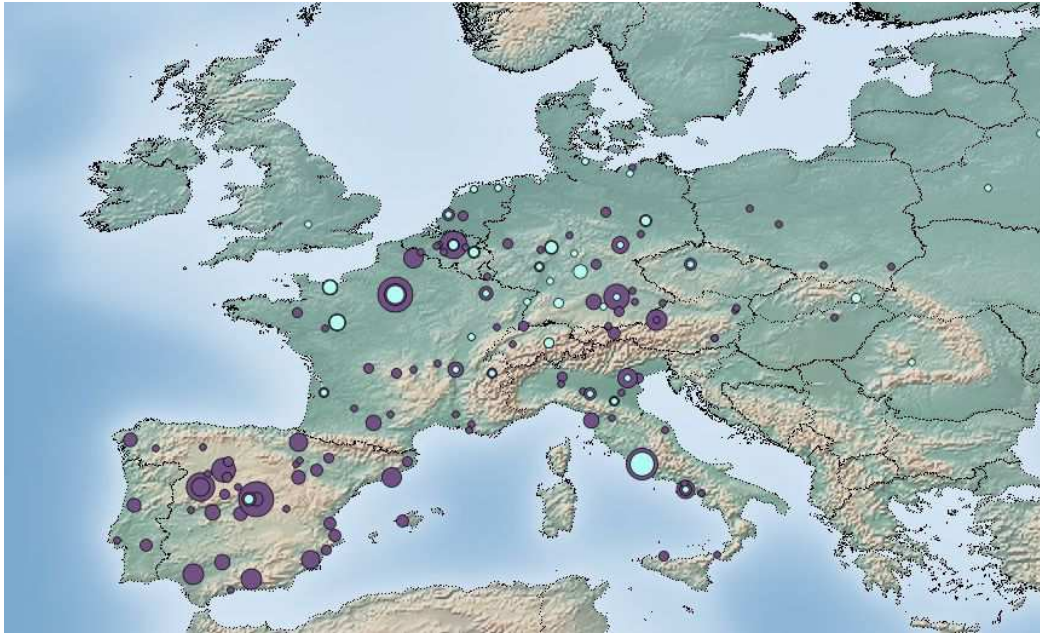
les caractéristiques essentielles des données, dont voici la distribution :



*Figure 13. Enseignements par discipline, regroupés et codés*

Le déséquilibre dans les données vient du fait que les notices se limitent souvent à indiquer que l'objet de l'enseignement est la théologie ou la philosophie, sans plus. Pour d'autres acteurs, on fait état de l'orientation ou du courant représenté, par exemple le thomisme ou le scotisme ; ces enseignements ont été regroupés sous l'appellation de disciplines théologiques (figure 13 : theol\_disc). Les lettres classiques, l'éloquence, le grec ont été classés en tant que lettres anciennes (lettr\_ant) alors que médecine théorique, la médecine et l'anatomie se retrouvent dans le groupe des sciences médicales. Et ainsi de suite. Enfin, si l'on s'interroge sur l'émergence de la nouvelle science, on peut regrouper des disciplines telles que les mathématiques, la physique expérimentale, l'astronomie, l'hydrographie sous le code « nouv\_science ». Tout en sachant qu'il faudrait faire varier les regroupements, comme le montre le cas du cartésianisme gardé à part dans notre codage, pour tester d'autres hypothèses, ou alors procéder à des classement plus fins car un enseignant de mathématiques n'est pas nécessairement un adepte de la nouvelle science. Toutefois, les informations de départ sont trop imprécises et elles ne se prêtent donc pas, en l'état, à une analyse de ce type.

En profitant de la géolocalisation de la plupart des enseignements, on peut s'interroger sur leur distribution dans l'espace. Nous avons donc procédé à des visualisations selon différents critères.



*Figure 14. Localisation et effectifs des enseignements de la philosophie  
et de la 'nouvelle science'*

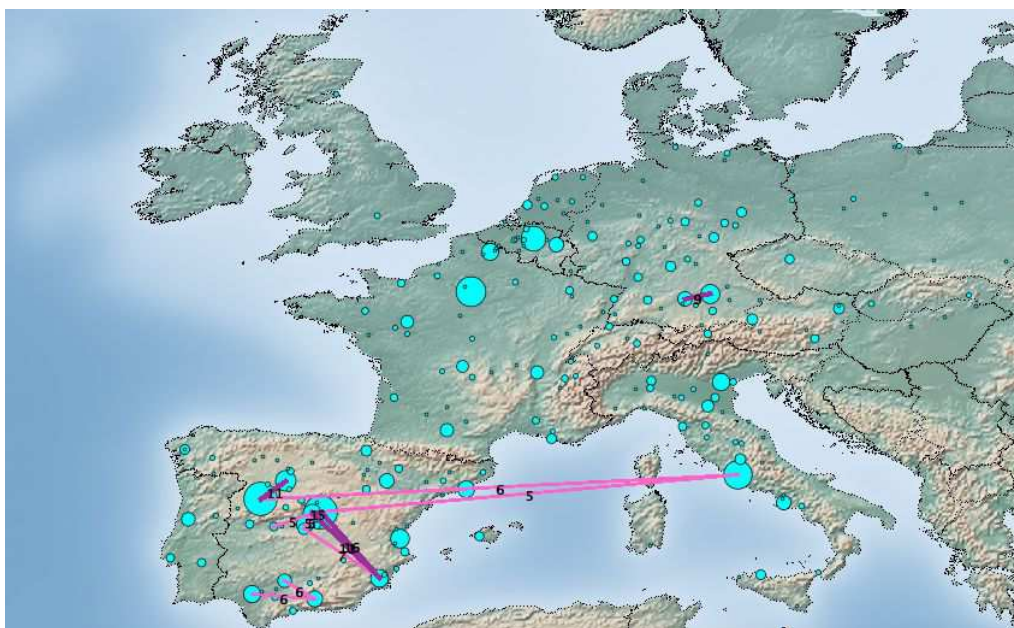
En brun figurent les effectifs des enseignements de la philosophie et en turquoise ceux de la nouvelle science (figure 14). Si l'on disposait de données plus précises sur les dates des enseignements, il serait possible de réaliser des cartes pour différentes périodes et visualiser l'évolution de l'enseignement de certaines disciplines. En l'état actuel, la carte permet de constater que les disciplines regroupées sous l'appellation de 'nouvelle science' sont bien représentées en France et en Allemagne, peu en Espagne et, en revanche, qu'elles ont l'effectif le plus important à Rome. Cet constat surprenant s'explique sans doute par la part prépondérante qui revient aux savants jésuites dans la population, ce qui provoque une sorte de surreprésentation des disciplines mathématiques dans la capitale des Etats pontificaux. En d'autres termes, si l'on considère les limites de la population exposées ci-dessus, il apparaît clairement que la carte des enseignements ne propose pas une représentation de la réalité de l'époque mais uniquement l'état de la population concernée, fortement marquée par la présence de membres de la Compagnie de Jésus.

On peut aussi comparer la distribution géographique des lieux de naissance avec celle des lieux d'enseignement par rapport à différentes sous-populations, ce qui amène à constater des différences intéressantes. Par exemple, les acteurs nés en Italie ou en France

(au sens de régions géographiques ayant par approximation les contours des États actuels), enseignent pour la plupart dans leur région d'origine, alors que les Espagnols se retrouvent en Espagne, en Italie, en Amérique latine et aux Philippines. En affinant l'analyse et en opérant une distinction entre les enseignants nés en Aragon et ceux nés en Castille, on constate que les premiers enseignent principalement dans la péninsule ibérique, alors que c'est le lot des deuxièmes de 'migrer' pour enseigner à l'échelle planétaire.

Quant aux membres des différents ordres religieux, on constate là aussi des configurations intéressantes : les dominicains ont leurs origines et leurs centres d'enseignement plutôt dans l'Europe du sud, ou alors dans les lieux universitaires traditionnels, tel Paris ou Louvain ; les clercs séculiers sont plutôt originaires de la France du Nord et des Pays-Bas mais ils se répandent pour enseigner dans toute l'Europe, et en particulier dans la péninsule ibérique ; quant aux jésuites, ils sont issus de l'ensemble de l'Europe, y compris centrale et orientale, et ils enseignent à l'échelle de la planète. C'est peut-être par rapport à cette question que la qualité des données, même si elle présente les limites que nous avons indiquées, permet de visualiser un phénomène structurel qui dépasse les limites de la population retenue.

On peut aussi s'interroger sur les migrations de chaque enseignant entre différentes institutions et différents lieux d'enseignement. On peut imaginer, en appliquant une méthode propre à l'analyse des réseaux, qu'un enseignant établit un lien entre deux institutions : on pourra ensuite supprimer les enseignants des données pour ne garder que les liens. On peut aussi faire 'disparaître' les institutions en les regroupant par lieu, tout en sachant que si on disposait de données de meilleure qualité, on pourrait proposer des analyses plus fines, distinguant entre les différentes institutions, ordres religieux d'appartenance, etc.

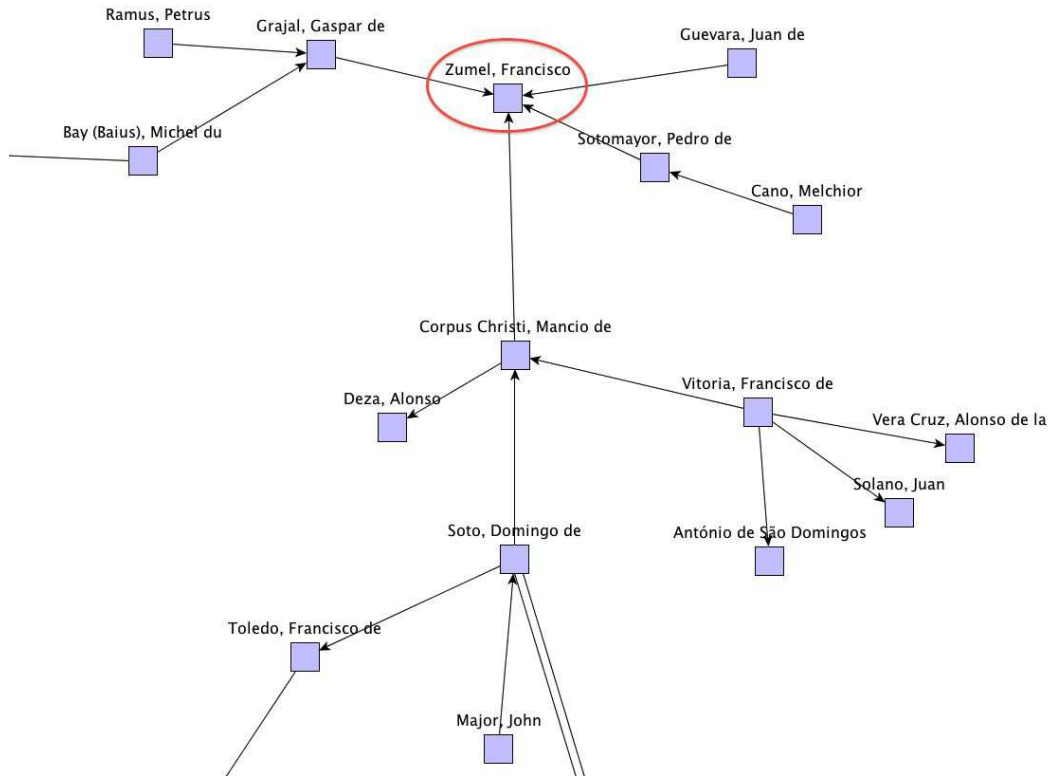


*Figure 15. Flux d'enseignants entre lieux d'enseignement*

En représentant ces liens sur une carte, on obtient une maille inextricable. Pour la démêler, on peut compter le nombre de liens par ligne et ne représenter que les lignes qui dépassent un seuil, par exemple, de cinq relations entre deux lieux. Les traits concernant des flux de plus de dix liens sont en couleur plus foncé (figure 15). Les points en turquoise représentent les effectifs des enseignements regroupés pour chaque lieu. Apparaît ainsi une certaine 'stabilité' des enseignants par rapport aux effectifs : seuls quelques flux significatifs sont visibles entre Rome, Alcalá et Salamanque, d'une part, ainsi qu'entre des régions spécifiques en Espagne, d'autre part, auxquels s'ajoutent les liens entre les collèges jésuites de Dillingen et Ingolstadt, dans le sud de l'Allemagne. Certes, si on baisse le seuil on verra apparaître plus de liens sur la carte mais on peut postuler que, sur une période de deux siècles, il faut un minimum de cinq liens pour déceler des relations significatives entre institutions. De nouveau, le problème est plutôt en amont, au niveau des informations que contiennent les notices, qui pourraient être enrichies et permettre une étude fort intéressante à ce sujet.

Enfin, un dernier exemple d'exploration possible est celui des filiations intellectuelles. Lors de la saisie d'informations de type 'Formation', il a été possible dans certains cas d'indiquer non seulement l'institution fréquentée mais également le maître auprès duquel l'acteur s'est formé. Hélas, seulement 115 relations de ce type ont été

saisies à partir du contenu du site web et, de plus, leur visualisation dans un logiciel d'analyse des réseaux<sup>12</sup> montre que dans la plupart des cas il ne s'agit que de relations entre deux individus.



*Figure 16. Filiations entre enseignants*

Mais la portion de ce graphe retenue dans la figure 16 illustre tout l'intérêt de cette approche car des filiations intéressantes apparaissent entre théologiens dominicains qui conduisent (le sens des flèches allant de l'enseignant vers l'élève) à Francisco Zumel (1540-1607), appartenant à l'Ordre des Mercédaire, qui interviendra dans la querelle *de auxiliis* en envoyant à Rome un « rapport complet sur les positions [du théologien jésuite] Luis de Molina »<sup>13</sup>.

<sup>12</sup> Pour une première exploration, j'ai utilisé le logiciel Visone, "<http://visone.info/>".

<sup>13</sup> "[http://scholasticon.fr/Database/Scholastiques\\_fr.php?ID=110](http://scholasticon.fr/Database/Scholastiques_fr.php?ID=110)".

## Conclusion

Ces exemples d'exploration des données produites par extraction d'informations structurées selon la sémantique de SyMoGIH à partir de *scolasticon.fr* montrent toute la richesse et tout le potentiel de ce site pour la recherche prosopographique. Ils en révèlent en même temps les limites, qui font que, en l'état actuel, ces données ne sont pas exploitables pour une analyse prosopographique.

Un premier problème est lié au caractère limité de la population retenue, en dépit du périmètre très large qui est visé par le site, s'étendant à l'ensemble des enseignants de philosophie et de théologie à l'époque moderne. Dans les faits, une large part revient aux membres espagnols de la Compagnie de Jésus et aux autres théologiens impliqués dans les débats intellectuels étudiés par J. Schmutz, créateur de *scholasticon.fr*. Deux stratégies sont envisageables à ce sujet. On peut opérer un rééquilibrage en élargissant de manière significative la population. Comme il n'est pas possible de réaliser ce projet en un seul coup pour trois siècles, on pourrait se concentrer sur une période précise, par exemple la première moitié du 17<sup>e</sup> siècle, pour laquelle le site comporte déjà une importante population. Ou alors, il serait possible d'assumer explicitement la spécialisation du site par rapport à une population ciblée, dont le périmètre serait lié aux compétences et aux intérêts des contributeurs, et d'en recenser systématiquement les individus afin que ces sous-populations d'enseignants soient complètes. On pourrait ensuite se mettre en réseau avec d'autres projets analogues, comme ceux portés par les chercheurs qui se fédèrent autour du groupe Héloïse<sup>14</sup>, pour échanger les données et les enrichir avec celles issues d'autres populations. A cette fin, on pourra tirer profit des technologies du web sémantique.

Un deuxième problème est lié au caractère incomplet et parfois imprécis de certaines notices. Nous avons vu que pour la moitié des acteurs manquent des indications sur leur enseignement suffisamment détaillées pour pouvoir en tirer des informations structurées. Il serait possible d'établir une liste des informations de base que l'on souhaite réunir pour chaque acteur recensé et de s'efforcer, du moins par approximation, de les fournir. Par exemple, il s'agirait de disposer des indications le plus précises possible concernant le lieu et l'année de naissance, ainsi que celles relatives aux formations et aux titres académiques acquis, et enfin les périodes d'enseignement et les matières enseignées. Il faudrait également élargir ces indications aux ouvrages publiés par ces acteurs. Tout en gardant une perspective bio-bibliographique, on disposerait ainsi d'un corpus de données permettant une première analyse systématique autour des questionnements présentés ci-

---

14 ["http://heloise.hypotheses.org/"](http://heloise.hypotheses.org/).



dessus.

Certes cette structuration n'est pas indispensable à la réalisation d'un site recensant des notices bio-bibliographiques et on pourrait la considérer trop coûteuse en temps et en recherches. Mais il me semble que cette impulsion venant d'une approche prosopographique amènerait les rédacteurs des notices à préciser la définition des objets mentionnés (les institutions concernées, les matières enseignées), en fournissant systématiquement quelques données de base pour chaque acteur, ce qui comporterait une qualification considérable de l'information mise à disposition. Cette démarche ne peut être réalisée que par les spécialistes du domaine et, en retour, elle permet de valoriser leur production de connaissances mise en œuvre pour la réalisation d'un tel site. Un élément essentiel de ce point de vue est représenté par un sourçage précis des informations, qui permet de les vérifier et de les mettre à jour progressivement.

Cette approche invite nécessairement à évoluer d'un système de notices–texte vers une base de données permettant de décomposer l'information et de la mobiliser depuis plusieurs points de vue, afin de produire un site web dynamique sur le modèle de sites existants et qui englobera également des cartes et des affichages analytiques des données. Etant donné le travail qui a été déjà accompli en appliquant la sémantique de SyMoGIH à ces informations, le coût de la migration de *scholasticon.fr* vers un système d'information géo-historique dynamique n'est pas très élevé. Cette migration aura l'avantage de promouvoir la démarche collaborative entre les spécialistes du domaine susceptibles d'alimenter directement la base de données. Pour ce qui concerne l'élargissement de la population ainsi que la recolte de nouvelles informations, il faudra prévoir la mise en place d'un projet disposant d'un financement adéquant. En sachant que si 9000 informations ont été produites en trois mois à partir des notices existantes, le dépouillement systématique des ouvrages bio-bibliographiques classiques, telles les bibliothèques des *scriptores* des ordres religieux, et de la bibliographie plus récente permettrait au cours de la durée habituelle de trois ans d'un projet financé de recollecter plusieurs dizaines de milliers d'informations. On disposera alors d'un volume et d'une qualité des données suffisante pour entamer une analyse, dans une approche prosopographique mais pas uniquement, de la « culture scolastique moderne ».

Francesco Beretta  
(CNRS LARHRA UMR 5190 – Université de Lyon)